Harding, Richard J.; Weedon, Graham P.; van Lanen, Henny A.J.; Clark, Douglas B.. 2014. **The future for global water assessment.**

The future for Global Water Assessment

Richard J. Harding[1], Graham P. Weedon[2], Henny A.J. van Lanen[3] and Douglas B. Clark[1].

1. Centre for Ecology and Hydrology, Wallingford, Oxon., United Kingdom.
2. Met Office Hadley Centre, Joint Centre for Hydrometerological Research, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK
3. Hydrology and Quantitative Water Management Group, Wageningen University, Netherlands

To be submitted to: *Journal of Hydrology*

Abstract

The global water cycle is a fundamental component of our climate and Earth system. Many, if not the majority, of the impacts of climate change are water related. We have an imperfect description and understanding of components of the water cycle. This arises from an incomplete observation of some of the stores and fluxes in the water cycle (in particular: precipitation, evaporation, soil moisture and groundwater), problems with the simulation of precipitation by global climate models and the wide diversity of global hydrological models currently in use. This paper discusses these sources of errors and, in particular, explores the errors and advantages of bias correcting climate model outputs for hydrological models using a single large catchment as an example (the Rhine). One conclusion from this analysis is that bias correction is necessary and has an impact on the mean flows and their seasonal cycle. However choice of hydrological model has an equal, if not larger effect on the quality of the simulation. The paper highlights the importance of improving hydrological models, which run at a continental and global scale, and the importance of quantifying uncertainties in impact studies.

Key Words

Water cycle, global, evaporation, river discharge, climate change, climate models

1. Introduction

The terrestrial water budget is at the heart of many environmental issues. Water is crucial to agricultural production, carbon budgets (and other biogeochemical cycles), biodiversity, energy generation, industrial production and human health. Extremes play an important role – floods and droughts are pressure points on water scarcity and environmental damage. There is increasing pressures on available water in many regions of the world due to increasing water demand because of a growing population and wealth and this is before the potential impacts of climate change. It is clearly important to develop well founded estimates of future water availability, as well as extremes, to underpin adaptation plans for the future.

Floods, droughts, increased water scarcity, reduced food and energy production – many of the key impacts of climate change identified by the IPCC are water related (Bates et al., 2008; IPCC 2007a). From the thermodynamics of the atmosphere we know increasing greenhouse gases are likely to lead to an increase in temperature, and this is already observed. Higher temperatures will increase evaporation, over the oceans in particular and hence water vapour in the atmosphere. This is likely to lead to overall higher rainfall globally and the likelihood of more intense rainfall regionally. Increases in rainfall intensities have been observed in more studies than decreases, although there are wide regional and seasonal variations (Berg et al., 2009; IPCC, 2012; Donat et al., 2012) and there are large areas of the world where there are not sufficiently long records of daily rainfall available to analyse.

Climate models continue to suggest increases of rainfall in the northern hemisphere high latitudes and decreases of rainfall in the sub-tropical (generally semi-arid) regions of the world, such as the Mediterranean, southern USA and Central America, south Australia and southern Africa (IPCC, 2007b). In other words wet areas get wetter and dry areas drier. When translated into river flows and available water, future water scarcity is likely to occur in these latter regions but is also likely in China, India and the Middle East, where populations and water consumption are rising fast (e.g. Hagemann et al., 2013; Gerten et al., 2011). The regional details of these changes are, however, very uncertain. The future response of river flows (and hence floods and droughts) at the basin scale will depend not only on the projected changes in rainfall patterns as determined by atmospheric circulation patterns, which are not always well represented in the climate models, but also on the regional-scale basin characteristics (e.g. physiography, land cover, geology, Laize et al.,

2010) and human interventions such as dams and water abstraction and irrigation (e.g. Haddeland et al., 2006; Adam et al., 2007; López-Moreno et al., 2009; Biemans et al., 2011).

As yet it is difficult to discern an increase in rainfall globally, partly because changes in precipitation in different regions tend to cancel out. There is evidence of increasing precipitation at high latitudes, decreasing precipitation in the subtropical regions and possibly changing distribution of precipitation in the tropics by the shifting position of the Intertropical Convergence Zone (see e.g. Zhang et al 2007). But the regional details of these changes remain very uncertain. There is good evidence that the extremes of rainfall have increased in Europe and worldwide (e.g. Klein Tank and Können, 2003; Zolina et al., 2010, Groisman et al., 2005, IPCC, 2012, Donat et al., 2012). Pal et al., (2011) has been able to conclude that the intense rainfall and floods in the UK in 2000 were significantly more likely due to increased greenhouse gases.

Many of the observed trends in the hydrological cycle can be attributed to human activities, but not necessarily to increases in greenhouse gases alone. Wu et al. (2013) argued that both changing greenhouse gases and regionally varying atmospheric aerosol loading has already affected the hydrological cycle. A general decrease in groundwater across the sub-tropics (for example in India, e.g. Rodell et al., 2009; Tiwari et al., 2009 and the mid-west of the USA Rodell et al., 2006) has been observed directly or inferred from GRACE satellite data and is almost certainly due to over extraction for irrigation. Analysis with a global hydrological model shows that in the sub-humid to arid areas the total global groundwater depletion has increased from 126 in 1960 to 283 km^3/yr in 2000 (Wada et al., 2010). The latter equals 39% of the global yearly groundwater abstraction. Gleeson et al. (2012) compared the rate of global groundwater depletion against the rate of natural renewal and the supply needed to support ecosystems. They illustrate that humans are overexploiting groundwater in many large aquifers that are critical to agriculture, especially in Asia and North America.

Terrestrial evaporation has increased through the 1980s and 90s, most probably due to decreasing aerosols (Jung et al., 2011). Since 2000 this increase may have levelled off as evaporation becomes increasingly controlled by soil water limitations rather than the energy available for evaporation. Increasing runoff and decreasing low summer flows in hundreds of near-natural catchments have been observed in Europe (Stahl et al., 2010). Flows in the northern rivers have increased (Peterson et al., 2002), but it is unclear whether this is due to land-cover change, increasing precipitation or indirect effects on water loss from plant transpiration linked to increasing CO_2 levels (see Gerten et al., 2008, Gedney et al., 2006). Gedney et al. (submitted) attributed long-term changes in discharge from large European basins to the combined effects of changing aerosol loading on solar radiation and CO_2 levels on stomatal closure.

It is very likely that global warming has influenced river flows, but often either the long-term river-flow data are not available or the natural changes are masked by anthropogenically-driven changes in land cover or water extraction. To reliably assess future water resources collaboration between climate, hydrological and water resource scientists working across a wide variety of scales is thus essential. In recent years considerable advances have been made with the bringing together of a wide variety of data sets and models (see e.g. Weedon et al., 2011, Haddeland et al., 2011, Harding et al., 2011).

2. Global Data Availability

There have been a number of initiatives to collate global precipitation data sets into gridded fields, for example Biemans et al., (2009) identifies seven such datasets. These vary with time step (monthly or daily), time period and spatial resolution. Although some of these datasets incorporate satellite data (for example the Global Precipitation Climatology Project, http://www.gewex.org/gpcp.html) and weather forecast analyses (CMAP, Xie and Arkin, 1997) all ultimately depend heavily on ground based rain gauge data. While some regions have dense rain gauge networks there are many regions, such as north and central Africa and the high latitudes, where networks are too sparse or the records not sufficiently continuous to provide a thorough assessment of trends and variability (e.g. Groisman et al., 2005). Mountainous regions also present considerable challenges; the networks being inevitably sparse and also biased towards low altitudes. Corrections have only recently been derived for precipitation gauges in mountainous areas (Adam et al., 2006) although the density of gauges and understanding of spatial variations of rainfall in mountainous regions remains inadequate.

River discharge is monitored widely around the world. The Global Runoff Data Centre (GRDC, http://www.bafg.de/) archives discharge data for almost 9000 gauging stations worldwide, two-thirds of which have daily data. However like the rainfall data the spatial (and temporal) coverage is patchy with large gaps in Africa (excluding South Africa) and Southern Asia. A new dataset of daily streamflow records for 10 countries across Europe, based on an updated version of the UNESCO FRIEND-Water European Water Archive (EWA), is useful for validating model outputs, — (Stahl et al., 2010; 2012, Hannaford et al., 2013). The dataset comprises catchments with minimal anthropogenic disturbances on flow regimes, monitored by gauging stations regarded to have good hydrometric performance with records from 1961 to 2005. The total dataset consists of 579 gauging stations. The distribution of stations over Europe is somewhat uneven with high densities of stations in some areas (e.g., Germany) and limited data in areas that are heavily affected by anthropogenic disturbances (e.g., northern France and the Benelux countries). No data were available across the majority of southern or eastern European countries. The availability of these key river measurements is hampered by the diversity of responsible organisations, a lack of investment and of the political will to share data.

Evaporation is a difficult quantity to measure and varies with land cover and soil types as well as climate so it is difficult to generalise. There are thus few routine measurements. The nearest we have to a global network is the FLUXNET data set (http://fluxnet.ornl.gov/, Baldocchi, 2008). This network consists of over 700 stations but not all the data are available and they cover variable (often short) time periods and are of variable quality. There have been a number of attempts to produce a gridded evaporation product – either based on the FLUXNET data and/or satellite retrievals (e.g. Miralles et al., 2011; Mueller at al., 2013). All these estimates depend on a model to derive evaporation from satellite products and/or meteorology or to extrapolate point measurements spatially and temporally. The mean of the global estimates is 1.56 mm d^{-1} (570 mm yr^{-1}) with a standard deviation of 0.2 mm d^{-1}, or just over 10 percent, regionally the spread can be much larger, of the order of 50% for large basins (Mueller et al., 2011).

Similarly there are limited networks of long-term in situ soil moisture measurements, although there are some notable exceptions, for example in the USA, Russia and China (see e.g. Entin et al., 2000). Increasing interest in soil moisture has led to some new measurements and the establishment of the International Soil Moisture Network (http://www.ipf.tuwien.ac.at/insitu/, Dorigo et al 2011). An interesting new development is the Cosmic ray soil moisture observing system (COSMOS, Zreda et al., 2012). COSMOS sensors, based on passive detection of scattered cosmic ray neutrons, have the advantage that they average at the field scale, approximately 600m diameter, thus removing much of the smallest-scale spatial variability. The measurement is also non-intrusive, automatic and does not require an internal neutron source for calibration. Despite this increasing activity there are still many regions of the land surface with no soil moisture measurements and even in those where there are the current network is often inadequate to provide representative regional figures.

Increasingly satellite products of soil moisture are available (AMSR-E SSM/i , SMOS etc, e.g. Loew et al., 2013;) and more are planned (such as ESA's Sentinel 1) however they monitor only the top few centimetres of the soil rather than the entire soil depth and are often strongly affected by thick vegetation cover. The GRACE (Gravity Recovery And Climate Experiment) satellites provide a unique data set which can retrieve an estimate of the total water storage (groundwater, soil water, snow cover and ice) though at low spatial resolution compared to the microwave satellite sensors. Since 2002 GRACE has provided considerable insights into changing regional ground water levels and seasonal variations of soil water (e.g. Rodell et al., 2009; Famiglietti et al., 2011; Houborg et al., 2012). The solution to these various incomplete measurement systems is almost certainly a data assimilation system combining measurements of different scales with one (or possibly an ensemble) of soil moisture models.

We therefore have incomplete measurements of water (and energy) budgets at country, continental and global scales. Given this lack of directly measured components of the water cycle the only way to obtain globally consistent estimates of the stores and fluxes is via hydrological modelling – informed and validated where possible with observations.

3. Uncertainty in estimates of the global terrestrial water budget.

In the last few years the Global Water System Project (GWSP) and the EU funded WATCH project (e.g. Harding et al., 2011) have co-ordinated an inter-comparison of hydrological models globally (WaterMIP, Haddeland et al., 2011). The inter-comparison has made use of a new global data set of meteorological data (the WATCH Forcing Data, WFD, Weedon et al., 2011). This is a combination of reanalysis products (ERA40) with observations (CRU TS2.1 and GPCCv4), thus the models used consistent driving data and a consistent terrestrial grid including a common river routing network. Eleven models were included in the intercomparison, including Global Hydrological Models and stand alone versions of the land surface models commonly used in climate models (Haddeland et al., 2011). The main distinction between these two classes of models is that Global Hydrological Models solve the water balance alone whereas the land surface models solve the energy and water balances (and often have a carbon budget). All but one of the models (WATERGAP) was run without calibration via observed discharge data. The initial analysis was for "naturalised" conditions (Haddeland et al., 2011) - i.e. excluding human influences related to land cover changes, damming, water abstraction and irrigation. Importantly, by concentrating on the late twentieth century, the performance of the hydrological models could be evaluated against observed basin discharge records.

The eleven models in WaterMIP showed a significant spread of the partitioning of precipitation into evapotranspiration and runoff. Averaged over the terrestrial surface (excluding Greenland and Antarctica) the average annual global evapotranspiration varied between models from 415 to 586 mm yr^{-1} and runoff from 290 to 457 mm yr^{-1}. There was no single cause for the spread in model outputs, although the different model treatment of snow was a major factor explaining the different shapes of the simulated annual hydrographs. Most models overestimate total annual runoff in semi arid regions – probably a result of both water extractions not being included in this phase of WaterMIP, and wetland evaporation, typically not being included in these models. Interestingly the runoff for the Brahmaputra was under-estimated – this is probably a result of the underestimate of precipitation in the Himalayan region.

Hannaford et al. (2010); Prudhomme et al. (2011), Gudmundsson et al. (2011; 2012a; 2012b), Van Loon et al. (2012), Van Huijgevoort et al. (2013) extended the analysis for a subset of the WaterMIP models to investigate one or both hydrological extremes (floods and droughts). Models were inter-compared, and compared against a precipitation index and against the European streamflow dataset of Stahl et al. (2010; 2012). The analyses concluded that the models generally identify the most extreme events and broadly show the same spatial-temporal resolution evolution of hydrological extremes, but variations in the representation of sub-surface flows and storage between models produce large variations in the simulated dynamics. All models struggle to reproduce the high observed flows –most probably because of the low spatial resolution of the input data (0.5 x 0.5^{o} or about 50 x 50 km) and have even more difficulties to simulate low flows.

4. Prediction of future flows

In order for future impacts of climate change to be assessed correctly it is essential that driving data are as realistic as possible. Current GCMs have substantial biases in their rainfall simulations. Most models overestimate precipitation, particularly over areas of complex topography and underestimate high intensity precipitation (see e.g. Mehran et al., 2012). For example for ECHAM6 model overall the precipitation is overestimated by 10%, with up to 5 mm day^{-1} in the tropics and 2 mm day^{-1} in mid latitudes (Stevens et al., 2012). In fact the errors in GCM daily precipitation are evident in the entire intensity spectrum, with too much low intensity drizzle and an underestimate of high precipitation events (see e.g. Piani et al., 2010b). Hydrological models involve thresholds and other non-linearities which result in incorrect trends and incorrect changes in extremes given the wrong input data. Most studies, therefore, use off line calculations of runoff flowing some degree of bias correction of the original GCM output (e.g. Hempel et al., 2013). This procedure has the added advantage of allowing the intercomparison of multiple climate and hydrological model combinations, thus providing an estimate of uncertainty in the hydrological sub-models. It has the disadvantage of neglecting and feedbacks between the land surface and atmosphere (Dadson et al., 2013) and introduces and inconsistencies between the land surface model of the GCM and the hydrological model. Given that it is unlikely that the biases in GCM outputs will be substantially reduced in the near future some sort of bias correction is inevitable if realistic driving data are to be provided to the hydrological models (Allen and Sollen, 2008; Lenderlink and Van Meijgaard, 2008).

The WaterMIP process (Figure 1) was the first stage in a comprehensive multimodel analysis of the 20th and 21st C terrestrial water cycle. The initial runs for the 21st C have perforce been made with the outputs from the AR4 runs (Hagemann et al., 2013) for which only a limited set of GCMs stored outputs suitable to run the full set of hydrological models. All climate models have biases in their precipitation (and other fields) – these biases are in both the mean and day-to-day variability. These biases have a substantial impact on runoff, when translated through the modelling chain (Sharma et al., 2007, Hansen et al., 2006 and Hagemann et al., 2011). Within the multi-model impacts framework described in Fig. 1, prior to the hydrological model runs a statistical bias correction of the GCM driving data based on quantile mapping was used to adjust both daily precipitation and daily temperature (Piani et al., 2010a; 2010b). Recently the WaterMIP study has been extended using the CMIP5 runs for the twenty first century (Taylor et al., 2012) within the ISI-MIP (Schiermeier, 2012; Dankers et al., 2014; Prudhomme et al., 2014).

The bias correction methodology mentioned here has many advantages – for example it corrects both the mean and the variability. However, by breaking the daily correlations between temperature and rainfall and failing to bias correct other associated hydrological drivers – such as humidity and radiation (Haddeland et al., 2012) - additional errors can therefore be introduced during bias correction. This methodology also fails to correct the sequencing of wet and dry days within a month which may be critical in determining the probability of floods and droughts (Zolina et al., 2010). Thus the introduction of bias correction may be necessary, given the current state of climate simulations globally and regionally, but does introduce additional uncertainties in estimates of future climate impacts (see e.g. Ehret et al., 2012).

5. *Influence of bias correction*

A number of studies have described the use of bias corrected GCM output to simulate river flows (Haerter et al., 2011: Hagemann et al., 2011; Chen et al., 2011). Below we explore the impact of a bias correction by comparing model outputs based on both bias-corrected and uncorrected forcing. We use a single large catchment (the Rhine) using two hydrological/land surface models and the outputs from a single climate model (IPSL). The model outputs for 1960-2001 are compared with daily observed naturalised discharge data.

The two land surface/hydrology models are:

a) JULES – the land surface model of the UKMO climate model (Best et al., 2011) simulates the water and energy budgets of the land surface at a sub-hourly time step. Evaporation is modelled using a modified Penman/Monteith equation coupled to a photosynthesis/surface conductance model. A multi-layer soil model generates surface and sub-surface runoff which is then routed through the river network using a linear routing model (Oki et al., 1999).

b) The Simple Synthetic Hydrology Model (SSHM) - is a transient soil-water balance model that is combined with a conceptual groundwater model. It uses daily precipitation, temperature and reference evaporation as forcing data, and it was applied to simulate time series of daily snow melt, snow storage, actual evapotranspiration, soil moisture storage, groundwater recharge and discharge (Van Lanen *et al*., 1996; 2013). Land use and soil data characterise the physical catchment structure. SSHM is based upon the FAO approach to compute actual evapotranspiration (Allen et al., 1998) and the widely-used HBV model (e.g. Seibert et al., 2000).

The models were run for 1960 to 2000 using climate model output (1960 to 2000) that was bias corrected using a technique which corrects the mean and probability distributions of the daily average air temperature and precipitation (Piani et al. 2010a; 2010b). The outputs are compared with daily naturalised discharge data from the GRDC (Figures 2 and 3).

i) Simulation of Rhine discharge

The MBE (Mean Bias Error as the percentage difference of an average variable, e.g. average modelled discharge, from the observed average, Weedon et al., 2013) for precipitation shows that the IPSL-corrected precipitation matches the WATCH Forcing Data precipitation within the error margins (Fig. 3). Since the bias correction is based on the WFD this small MBE confirms that the bias correction was correctly applied. On the other hand, the original raw IPSL precipitation for 1960-2000 is close to double the WFD precipitation (MBE circa 100%, Fig. 3).

Comparison of the observed- and modelled-daily discharge MBE (Figure 3) provides an indication of the long-term (multi-year) water balance. The discharge MBE for both JULES-WFD and SSHM-WFD are significantly positive compared to the observed discharge indicating that both models discharge too much water overall – this is particularly true of SSHM and is presumably due to inadequacies in the evaporation formulations (too little net evaporation annually). Use of the corrected IPSL forcing for both JULES and SSHM substantially improved discharge MBE compared to using the WFD forcing. The overall corrected IPSL precipitation is essentially the same as the WFD precipitation. Thus changes in discharge MBE relate to differences between other atmospheric forcing variables such as net radiation, wind speed and humidity that are not affected by the bias correction method of Piani et al.(2010a and b).

The discharge MBE for JULES forced with the un-corrected IPSL data is larger than for forcing with either the WFD or the corrected-IPSL data. A component of this over-estimation must be the excessive un-corrected IPSL precipitation (indicated by the precipitation MBE). By contrast, SSHM discharge MBE is slightly under-estimated given the uncorrected IPSL forcing rather than over-estimated for the other forcing. As the uncorrected IPSL precipitation is about double both the WFD precipitation and corrected IPSL precipitation, the underestimation of SSHM discharge when forced with uncorrected IPSL data implies too much modelled evaporation.

ii) Annual cycles in Rhine discharge.

Spectrally the largest identifiable component of the daily discharge variability is the strong annual cycle (observed or modelled with any forcing). Focussing discussion on the annual scale makes it easier to interpret the reasons for differences between model output and observations and between different forcings. The amplitude-ratio and phase (or timing) of modelled discharge at the annual scale is used for comparison with the GRDC naturalised discharge observations (for methodology see Weedon et al., submitted). Note that for the WFD-forced model output the short-term (sub-annual) variability should ideally match the GRDC record hence the observed discharge (red) is plotted on top of the JULES-WFD discharge (black) and SSHM-WFD discharge (black, Fig. 2a). However, for the IPSL data derived from a GCM run, the specific meteorological evolution is not expected to match the actual meteorological history of 1960-2000. Thus the comparison is restricted to comparing the average characteristics of the annual cycles in modelled discharge with the annual cycles in the observed discharge (Figs 2b and c and 3 b and c).

Forced with the WFD, JULES has an annual cycle in discharge that is too large (Figs 2a and 3b), but with the correct timing (phase, Fig. 3c). While summer JULES-WFD and GRDC discharge are similar (Fig 2a) the JULES-WFD baseflow in winter is too large. This is probably because of the underestimation of overall evaporation (MBE about +20%).However, SSHM forced with the WFD has an annual cycle with approximately the correct amplitude and timing (within error) despite an MBE of about +60%. Thus SSHM-WFD correctly represents the variations (amplitude and phase) of the baseflow, but underestimates the overall evaporation so the MBE is too high.Visual inspection of Figure 2a shows that in terms of sub-annual discharge variations linked to precipitation events JULES-WFD reproduces the large sub-annual variability seen in the GRDC data pretty well. However, SSHM has only muted short-term discharge events compared to observations perhaps linked to limited surface runoff modelling.

Using the corrected IPSL forcing JULES has an annual cycle in discharge that is too large compared to observations and slightly larger than under WFD forcing, but still with the right timing (Fig. 3). Since the overall discharge in JULES-IPSL-corrected is lower than JULES-WFD (MBE about +10%), the larger amplitude annual cycles may reflect more evaporation in the summer compared to JULES-WFD - as supported by the lower average baseflow in the summer (Fig. 2b).

The SSHM annual discharge cycle is too large when forced by the IPSL-corrected data - and is much larger than for SSHM-WFD (Fig. 3b). As the discharge MBE is lower overall (about +10% versus +60%) the large annual cycles may reflect much larger amounts of evaporation in the summer as reflected by the substantially lower baseflow compared to the SSHM-WFD run.

JULES-not-corrected-IPSL has very large annual cycles in discharge (Figs. 2c and 3). This at least partly relates to the large uncorrected precipitation input as reflected by the much higher baseflow in winter than the other JULES runs. However, in addition summer baseflow is lower than for JULES-WFD and JULES-corrected-IPSL so apparently summer evaporation is larger than before - adding to the amplitude of the annual cycle.

The SSHM run with the uncorrected IPSL data has an annual cycle that appears to be too small in amplitude - though it is within error (95% confidence interval) of agreeing with observations. Since the discharge MBE is lower than expected there may be too much evaporation overall. However, the summer baseflow is very similar to the winter baseflow on average so the summer evaporation appears to be far too low.

The message from this analysis of the case study is that in order to obtain realistic discharge estimates, as judged against observations, bias correction is necessary and has an impact on the mean flows and their seasonal cycle. However choice of hydrological model has an equal, if not larger effect on the quality of the simulation. This conclusion supports that of Hagemann et al. (2013).

6. Conclusions

Water resources are already under considerable pressure in many parts of the world. These pressures will increase with global changes – particularly increasing population and affluence leading to increasing water extraction and land cover change. Climate change will also add to these pressures with dry areas getting drier and an increasing proportion of precipitation falling in extreme events, also leading to longer dry spells. Any adaptation measures must be strongly underpinned by a good knowledge of the current regime and an understanding of possible future regimes. This can only be obtained by a combination of models and data. Considerable progress has been made in recent years in the compilation of global data sets and in our understanding of model errors.

The substantial spread found between hydrological models commonly used for impact analysis suggests a single impact model should be used with great care (Haddeland et al., 2011; Stahl et al., 2012; Van Huijgevoort et al., 2013; Hagemann et al., 2013). This study also suggests that improvements to hydrological models used at large scales could and should be made, in particular in the high and low flow domain. Obvious examples are the improvement of evapotranspiration and snowfall components of models, to improve the overall water balance, and the inclusion of additional hydrological processes, such as ground water, permafrost and wetlands, to improve the low and high flow representation. The use of calibration via spatially-aggregated local observations is an additional aspect which should be carefully considered. Calibration can undoubtedly improve radically the simulation within a single basin; however, it can hide structural weaknesses within a particular model. It may also reduce the global applicability of a model – it is clear that the modelling suite used must be considered carefully against its purpose.

Global climate models still show persistent regional biases in precipitation and a tendency to produce too much light rain (see e.g. Perkins et al., 2007). Regional climate models fare a little better but still have considerable biases (e.g. Rawlins et al., 2012). Mean runoff can be seen as a residual of the precipitation after subtraction of the evaporation so any bias in the precipitation is likely to be amplified, in percentage terms, in the runoff. In addition many of the runoff processes are strongly non-linear, thus the variability (in time and space) is important. While simulations of precipitation are improving progress is slow and there is little prospect of dramatic advances in the next few years. In the meantime society needs regional and local assessments of the impact of climate change and this cannot wait for the advance of climate models. The need for the correction of biases in the GCM outputs to provide realistic estimate of runoff (and changes in runoff) is demonstrated in this paper. It is also clear that the current state-of-the-art bias-correction methodologies need further refinements. Hempel et al. (2013) have developed the bias correction methodology of Piani et al. (2010a) to include other variables, such as radiation, but problems with cross correlations still persist.

Estimations of the water cycle for the future contain many uncertainties – GHG scenarios, climate model uncertainties, hydrology/climate feedbacks, bias correction, imperfect large-scale hydrological models, water use/exploitation scenarios. We need a new framework for impact model assessment which should include: common driving data, common (and explicit) land use and extraction scenarios, ensembles of climate and hydrological models and uncertainty description. The new ISI-MIP approach is a useful step towards an integrated inter-sectorial approach in impact assessment (Piontek et al., 2014).

Acknowledgements.

This paper was developed as part of the European Union (EU-FP6)-funded integrated program called WATCH (Contract 036946). G.P.W. was supported by the Joint DECC and DEFRA Integrated Climate Programme DECC/DEFRA (GA011101). H.A.J.v.L was also supported by the EU-FP7 collaborative project DROUGHT-R&SPI (contract 282769). Niko Wanders (Utrecht University, the Netherlands) helped to retrieve and process data for SSHM.

References

Adam, J.C., E.A. Clark, D.P. Lettenmaier, and E.F. Woods, 2006. Correction of global precipitation for orographic effects. *J. Climate* 19, 15-38.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M., 1998. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*, FAO Irrig. Drain. paper 56, Rome, 1988.

Allen, R.P., B.J. Soden, 2008. Atmospheric warming and the amplification of precipitation extremes. *Science* 321, 1481-1484.

Baldocchi, D.D . 2008. 'Breathing' of the Terrestrial Biosphere: Lessons Learned from a Global Network of Carbon Dioxide Flux Measurement Systems. *Australian J. Botany* 56, 1-26.

Bates, B., Kundzewicz, Z., Wu, S., and Palutikof, J., 2008. Climate Change and Water. *Technical Paper of the Intergovernmental Panel on Climate Change, Intergovernmental Panel on Climate Change, Geneva, 2008.*

Berg, P., J.O. Haerter, P. Thejll, C. Piani, S. Hagemann, and J.H. Christensen, 2009. Seasonal characteristics of the relationship between daily precipitation intensity and surface temperature. *J. Geophys. Res.* 114, D18102, doi:10.1029/2009JD012008.

Best, M.J., M. Pryor, D.B. Clark, G.G. Rooney, R.L.H. Essery, C.B. Ménard, J.M. Edwards, M.A. Hendry, A. Porson, N. Gedney, L.M. Mercado, S. Sitch, E. Blyth, O. Boucher, P.M. Cox, C.S.B. Grimmond, and R.J. Harding, 2011. The Joint UK Land Environment Simulator (JULES), model description - Part 1: Energy and water fluxes. *Geosci. Model Dev.* 4, 677-699.

Biemans, H., I. Haddeland, P. Kabat, F. Ludwig, R.W.A. Hutjes, J. Heinke, and D. Gerten, 2011. Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resour. Res.* , 47, W03509, doi:10.1029/2009WR008929.

Chen, C., J.O. Haerter, S. Hagemann and C. Piani, 2011. On the contribution of statistical bias correction to the uncertainty in the projected hydrological cycle. *Geophys. Res. Lett.*, 38, L20403, doi:10.1029/2011GL049318.

Dadson, S., Acreman, M. and Harding, R. , 2013. Water security, global change and land-atmosphere feedbacks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(2002).

Dankers, R., N.W. Arnell, D.B. Clark, P.D. Falloon, B.M. Fekete, S.N. Gosling, J. Heinke, H. Kim, Y. Masaki, Y. Satoh, T. stacke, Y. wada, D. Wisser, 2014. First look at changes in flood hazard in the Inter-Sectoral Imact Model Intercomparison Project ensemble. *Proc. Nat. Acad. Sci.* doi:10.1073/pnas.1302078110.

Donat, M. G., et al., 2013. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. Atmos.*, 118, 2098–2118, doi:10.1002/jgrd.50150.

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., and Jackson, T., 2011. The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements, *Hydrol. Earth Syst. Sci., 15, 1675-1698, doi:10.5194/hess-15-1675-201.*

Ehret, U., E.Zehe, V. Wulfmeyer, K. Warrach-Sagi and J. Liebert, 2012. Should we apply bias correction to global and regional climate model data? *Hydrol. Earth Syst. Sci.* 16, 3391-3404, doi:10.5194/hess-16-3391-2012.

Entin, J.K., Robock A., Vinnikov, K.Y. S, Hollinger, S.E., Liu, S., and A . Namkhai. 2000, Temporal and spatial scales of observed soil moisture variations in the extratropics. *J. Geophys. Res.*, 105, 11,865-11,877.

Famiglietti, J., M. Lo, S.L. Ho, J. Bethune, K.J. Anderson, T.H. Syed, S.C. Swenson, C.R. de Linage, and M. Rodell, 2011. Satellites measure recent rates of groundwater depletion in California's central valley *Geophys. Res Letter*, *38,* L03403, 10.1029/201GL046442.

Gedney, N., Cox, P.M., Betts, R.A., Boucher, O., Huntingford, C. and Stott, P.A. 2006. Detection of a direct carbon dioxide effect in continental river runoff records. *Nature* 439: 835-838.

Gedney, N., C. Huntingford, G.P. Weedon, N. Bellouin, O. Boucher and P.M. Cox, submitted. Detection of solar dimming and brightening effects on northern hemisphere river flow. *Nature Geosci.*

Gerten, D.**,** Rost, S., von Bloh, W., Lucht, W. 2008. Causes of change in 20th century global river discharge. *Geophys. Res. Lett.* 35, L20405.

Gerten, D., Heinke, J., Hoff, H., Biemans, H., Fader, M., Waha, K., 2011. Global water availability and requirements for future food production. *J. Hydrometeor.* 12, 885–899.

Gleeson, T., Wada, Y., Bierkens, M.F.P. and Van Beek, L.P.H., 2012. Water balance of global aquifers revealed by groundwater footprint. *Nature* 488, 197-200, doi:10.1038/nature11295.

Groisman, P.Y., Knight, R.W., D. R. Easterling, T. R. Karl, G. C. Hegerl, and V. N. Razuvaev, 2005. Trends in intense precipitation in the climate record. *J. Climate*, 18, 1343–1367.

Gudmundsson, L., Tallaksen, L.M. & Stahl, K., 2011. Spatial cross-correlation patterns of European low, mean and high flows. *Hydrol. Proc.* 25, 1034 –1045, doi: 10.1002/hyp.7807.

Gudmundsson, L., L.M. Tallaksen, K. Stahl, D.B. Clark, E. Dumont, S. Hagemann, N. Bertrand, D. Gerten, J. Heinke, N. Hanasaki, F. Voß and S. Koirala, 2012a. Comparing large-scale hydrological models to observed runoff percentiles in Europe. *J. Hydrometeor.*, 13, doi: 10.1175/JHM-D-11-083.1: 604-620.

Gudmundsson, L., Wagener, T., Tallaksen, L.M. & Engeland, K., 2012b. Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe, *Water Resour. Res.*, 48: W11504, doi:10.1029/2011WR010911.

Haddeland, I., T. Skaugen, and D.P. Lettenmaier, 2006. Anthropogenic impacts on continental surface water fluxes. *Geophys. Res. Lett.* , 33(8), L08406, doi:10.1029/2006GL026047.

Haddeland, I., D.B. Clark, W. Franssen, F. Ludwig, F. Voß, N.W. Arnell, N. Bertrand, M. Best, S. Folwell, D. Gerten, S. Gomes, S.N. Gosling, S. Hagemann, N. Hanasaki, R. Harding, J. Heinke, P. Kabat, S. Koirala, T. Oki, J. Polcher, T. Stacke, P. Viterbo, G.P. Weedon, and P. Yeh, 2011. Multimodel Estimate of the Terrestrial Global Water Balance: Setup and First Results. *J. Hydrometeor.*, 12, 869–884, doi: 10.1175/2011JHM1324.1.

Haddeland, I., J. Heinke, F. Voβ, S. Eisner, C. Chen, S. Hagemann, and F. Ludwig, 2012. Effects of climate model radiation, humidity and wind estimates on hydrological simulations. *Hydrol. Earth Syst. Sci.* 16, 305-318.

Hagemann, S., C. Chen, J.O. Haerter, J. Heinke, D. Gerten and C. Piani, 2011. Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. *J. Hydrometeor.* 12, 10.1175/2011JHM1336.1: 556-578.

Hagemann, S., Chen, C., Clark, D. B., Folwell, S., Gosling, S. N., Haddeland, I., Hanasaki, N., Heinke, J., Ludwig, F., Voss, F., and Wiltshire, A. J., 2013. Climate change impact on available water resources obtained using multiple global climate and hydrology models, *Earth Syst. Dynam.*, 4, 129-144, doi:10.5194/esd-4-129-2013.

Hannaford, J., Lloyd-Hughes, B., Keef, C., Parry, S., and Prudhomme, C., 2010. Examining the large-scale spatial coherence of European drought using regional indicators of precipitation and streamflow deficit. *Hydrol. Process.*, 25, 1146–1162.

Hannaford, J., Buys, G., Stahl, K., and Tallaksen, L. M., 2013. The influence of decadal-scale variability on trends in long European streamflow records. *Hydrol. Earth Syst. Sci. Discuss.*, 10, 1859–1896, doi:10.5194/hessd-10-1859-2013.

Harding, R., Best, M., Blyth, E., Hagemann, S., Kabat., P., Tallaksen, L.M., Warnaars, T., Wiberg, D., Weedon, G.P., van Lanen, H.A.J., Ludwig, F. & Haddeland, I., 2011. Preface to the "Water and Global Change (WATCH) special collection: Current knowledge of the terrestrial Global Water Cycle". *J. Hydrometeorol.* 12(6): 1149-1156, DOI: 10.1175/JHM-D-11-024.1.

Haerter, J.O. , S. Hagemann, C. Moseley and C. Piani, 2011. Climate model bias correction and the role of timescales. *Hydrol. Earth Syst. Sci.* 15, doi:10.5194/hess-15-1-2011: 1065-1079.

Hansen, J. W., A. Challinor, A. Ines, T. Wheeler, and V. Moron, 2006. Translating climate forecasts into agricultural terms: Advances and challenges. *Climate Res.*, **33**, 27–41.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F., 2013. A trend-preserving bias correction – the ISI-MIP approach. *Earth Syst. Dynam. Discuss.*, 4, 49-92, doi:10.5194/esdd-4-49-2013, 2013.

Houborg, R., M. Rodell, B. Li, Reichle, R., and Zaitchik, B. F., 2012. Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations *Water Res. Res.*, *48* (W07525) 10.1029/2011WR011291.

IPCC, 2007a: *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on*

Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 976pp.

IPCC, 2007b: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.

IPCC, 2012: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation.* A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.

Jung, M. et al., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, 467 951–4.

Klein, A. M. G. and Können, G.P., 2003. Trends in Indices of Daily Temperature and Precipitation Extremes in Europe,1946–99. *J. Clim.* 16: 3665-3680.

Laizé, C. R. L. and Hannah, D. M., 2010. Modification of climate-river flow associations by basin properties. *J. Hydrol.*, 389, 186–204.

Lenderlink, G. and E. Van Meijgaard, 2008. Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geosci.* 1, 511-514.

Loew, A., T. Stacke, W. Dorigo, R. de Jeu, and S. Hagemann, 2013. Potential and limitations of multidecadal satellite soil moisture observations for climate model evaluation studies. *Hydrol. Earth Syst. Sci. Discuss.*, 10, 3541-3594.

López-Moreno, J. I., S. M. Vicente-Serrano, S. Beguerı´a, J. M. Garcı´a-Ruiz, M. M. Portela, and A. B. Almeida, 2009. Dam effects on droughts magnitude and duration in a transboundary basin: The Lower River Tagus, Spain and Portugal. *Water Resour. Res.*, 45, W02405, doi:10.1029/2008WR007198.

Mehran, A., AghaKouchak, A., and Phillips, T. J., 2014. Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations, *J. Geophys. Res. Atmos., 119, doi:10.1002/2013JD021152*

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J., 2011. Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.*, 15, 453-469, doi:10.5194/hess-15-453-2011.

Mueller, B., S. I. Seneviratne, C. Jiménez, T. Corti, M. Hirschi, G. Balsamo, P. Ciais, P. Dirmeyer, J. B. Fisher, Z. Guo, M. Jung, F. Maignan, M. F. McCabe, R. Reichle, M. Reichstein, M. Rodell, J. Sheffield, A. J. Teuling, K. Wang, E. F. Wood, and Y. Zhang, 2011. Evaluation of global observations based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett*., 38, L06402, doi:10.1029/2010GL046230.

Mueller, B., M. Hirschi, C. Jimenez, P. Cias, P.A. Dirmeyer, A.J. Dolman, J.B. Fisher, M. Jung, F. Ludwig, F. Maignan, D.G. Miralles, M.F. McCabe, M. Reichstein, J. Sheffield, K. Wang, E.F. Wood, Y. Zhang, and S.I. Seneviratne, 2013. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrol. Earth Syst. Sci.*, 17, 3707-3720, doi:10.5194/hess-17-3707-2013.

Oki, T., T. Nishimura and P. Dirmeyer, 1999. Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP). *J. Met. Soc. Japan*, *77: 235-255.*

Pall, P., T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. J. Hilberts, D. Lohmann, and M. R. Allen, 2011: Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature,* 470, 382–385, doi:10.1038/nature09762.

Perkins, S. E., A. J. Pitman, N. J. Holbrook, and J. McAneney, 2007. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature and precipitation over Australia using probability density functions. *J. Climate*, 20, 4356–4376.

Peterson, B. J., R. M. Holmes, J. W. McClelland, C. J. Vorosmarty, R. B. Lammers, A. I. Shiklomanov, I. A. Shiknomanov, and S. Rahmstorf, 2002. Increasing river discharge to the Arctic Ocean. *Science, 298, 2171– 2173.*

Piani, C., G.P. Weedon, M. Best, S.M. Gomes, P. Viterbo, S. Hagemann, and J. O. Haerter, 2010a. Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol.*, 395, doi:10.1016/j.jhydrol.2010.10.024, 199–215.

Piani, C., J. Haerter, and E. Coppola, 2010b. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoret. Appl. Climat.*, 99:187-192, doi:10.1007/s00704-009-0134-9.

Piontek, F., C. Müller, T.A.M. Pugh, D.B. Clark, D. Deryng, J. Elliott, F.de J.C. González, M. Flörke, C. Folberth, W. Franssen, K. Frieler, A.D. friend, S.N. Gosling, D. Hemming, N. Khabarov, H. Kim, M.R. Lomas, Y. Masaki, M. Mengel, A. Morse, K. Neumann, K. Nishina, S. Ostberg, R. Pavlick, A.C. Ruane, J. Schewe, E. Schmid, T. Stacke, Q. Tang, Z. Tassler, A.M. Tompkins, L. Warszawski, D. Wisser and H.J. Schellnhuber, 2014. Multisectoral climate impact hotspots in a warming world. *Proc. Nat. Acad. Sci.* doi: 10.1037/pnas.1222471110.

Prudhomme, C., Parry, S., Hannaford, J., Clark, D. B., Hagemann, S., and Voss, F., 2011. How well do large-scale models reproduce regional hydrological extremes in Europe? *J. Hydrometeorol.*, 12, 1181–1204, doi:10.1175/2011JHM1387.1.

Prudhomme, C., I. Giuntoli, E.L. Robinson, D.B. Clark, N.W. Arnell, R. Dankers, B.M. Fekete, W. franssen, D. Gerten, S.N. Gosling, S. Hagemann, D.M. Hannah, H. kim, Y. Masaki, Y. Satoh, T. Stacke, Y. Wada, D. Wisser, 2014. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proc. Nat. Acad. Sci.* doi:10.1073/pnas.1222473110.

Rawlins, M. A., R. S. Bradley, and H. F. Diaz, 2012. Assessment of regional climate model simulation estimates over the northeast United States. *J. Geophys. Res.*, 117, D23112, doi:10.1029/2012JD018137.

Rodell M, Chen J, Kato H, Famiglietti J, Nigro J, Wilson C., 2006. Estimating ground water storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeol. J.* doi:10.1007/s10040-006-0103-7.

Rodell, M., I. Velicogna, and J.S. Famiglietti, 2009. Satellite-based estimates of groundwater depletion in India. *Nature*, *460*, 999-1002, 10.1038/nature08238.

Schiermerier, Q., 2012. Models hone picture of climate impacts. *Nature, 482, 286.*

Seibert, J., Uhlenbrook, S., Leibundgut, C., and Halldin, S., 2000. Multiscale calibration and validation of a conceptual rainfall-runoff model. *Phys. Chem. Earth B,* 25, 59–64.

Sharma, D., A. Das Gupta, and M. S. Babel, 2007. Spatial disaggregation of bias corrected GCM precipitation for improved hydrologic simulation: Ping River basin, Thailand. *Hydrol. Earth Syst. Sci.*, **11**, 1373–1390.

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., Demuth, S., Fendekova, M., and Jódar, J., 2010. Streamflow trends in Europe:

evidence from a dataset of near-natural catchments, *Hydrol. Earth Syst. Sci.*, 14, 2367–2382, doi:10.5194/hess-14-2367-2010.

Stahl, K., Tallaksen, L. M., Hannaford, J., and Van Lanen, H. A.J., 2012. Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble. *Hydrol. Earth Syst. Sci.*, 16, 2035–2047, doi:10.5194/hess-16-2035-2012.

Stevens, B., et al., 2012. The atmospheric component of the MPI-M Earth System Model: ECHAM6. J. Adv. Model. Earth Syst., doi:10.1002/jame.20015.

Taylor K. E., Stouffer R. J. and Meehl G. A., 2012. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* 93**,** 485–98.

Tiwari, V.M., Wahr, J.M. and Swenson, S., 2009. Dwindling groundwater resources in northern Indian region, from satellite gravity observations. *Geophys. Res. Lett.*, 36, L18401, doi:10.1029/2009GL039401.

Van Huijgevoort, M., P. Hazenberg, H. Lanen, R. Teuling, D. Clark, S. Folwell, S. Gosling, N. Hanasaki, J. Heinke, S. Koirala, T. Stacke, F. Voss, J. Sheffield, and R. Uijlenhoet, 2013: Global multi-model analysis of drought in runoff for the second half of the 20th century. *J. Hydrometeor. doi:10.1175/JHM-D-12-0186.1,early press release.*

Van Lanen, H. A. J., Weerts, A. H., Kroon, T., and Dijksma, R., 1996. Estimation of groundwater recharge in areas with deep groundwater tables using transient groundwater flow modelling. *Proceedings International Conference on "Calibration and Reliability of Groundwater Modelling",* September 1999, Golden, USA, 307–316.

Van Lanen, H.A.J., Wanders, N., Tallaksen, L.M. & Loon, A.F. van, 2013. Hydrological drought across the world: impact of climate and physical catchment structure. *Hydrol. Earth Syst. Sci.* 17: 1715–1732, doi:10.5194/hess-17-1715-2013.

Van Loon, A. F., Van Huijgevoort, M. H. J., and Van Lanen, H.A. J., 2012. Evaluation of drought propagation in an ensemble mean of large-scale hydrological models. *Hydrol. Earth Syst. Sci.*, 16, 4057–4078, doi:10.5194/hess-16-4057-2012.

Wada, Y., Van Beek, L.P.H., Van Kempen, Ch.M, Reckman, J.W.T.M., Vasak, S., Bierkens, M.F.P., 2010. Global depletion of groundwater resources. *Geophys. Res. Lett.* 37, L20402, doi:10.1029/2010GL044571.

Weedon GP, Gomes S, Viterbo P, Shuttleworth WJ, Blyth E, Österle H, et al., 2011. Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeorol*. [doi: 10.1175/2011JHM1369.1].12, 823-48.

Weedon, G.P., Prudhomme, C., Crooks, S., Ellis, R.J., Folwell, S.S. and Best, M.J. submitted. Evaluating the performance of hydrological models via cross-spectral analysis: case study of the Thames Basin, UK. *J. Hydrometeorol.*

Wu, P., Christidis, N., and Stott, P., 2013. Anthropogenic impact on Earth's hydrological cycle. *Nature Climate Change* doi: 10.1038/NCLIMATE1932.

Zhang X, Zwiers FW, Hegerl GC, Lambert FH, Gillett NP, Solomon S, et al., 2007. Detection of human influence on twentieth-century precipitation trends. *Nature*, 448, 461-5.

Zolina, O., C. Simmer, S.K. Gulev and S. Kollet, 2010. Changing structure of European precipitation: longer wet periods leading to more abundant rainfalls. *Geophys. Res. Lett*, 37, L06704, doi: 10.1029/2010GL042468.

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., Rosolem, R., and Ferre, T. P. A., 2012. COSMOS: The Cosmic-ray Soil Moisture Observing System. *Hydrol. Earth Syst. Sci. Discuss.*, 9, 4505-4551, doi:10.5194/hessd-9-4505-2012.

Xie, P. and P. A. Arkin, 1997. Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539-2558.

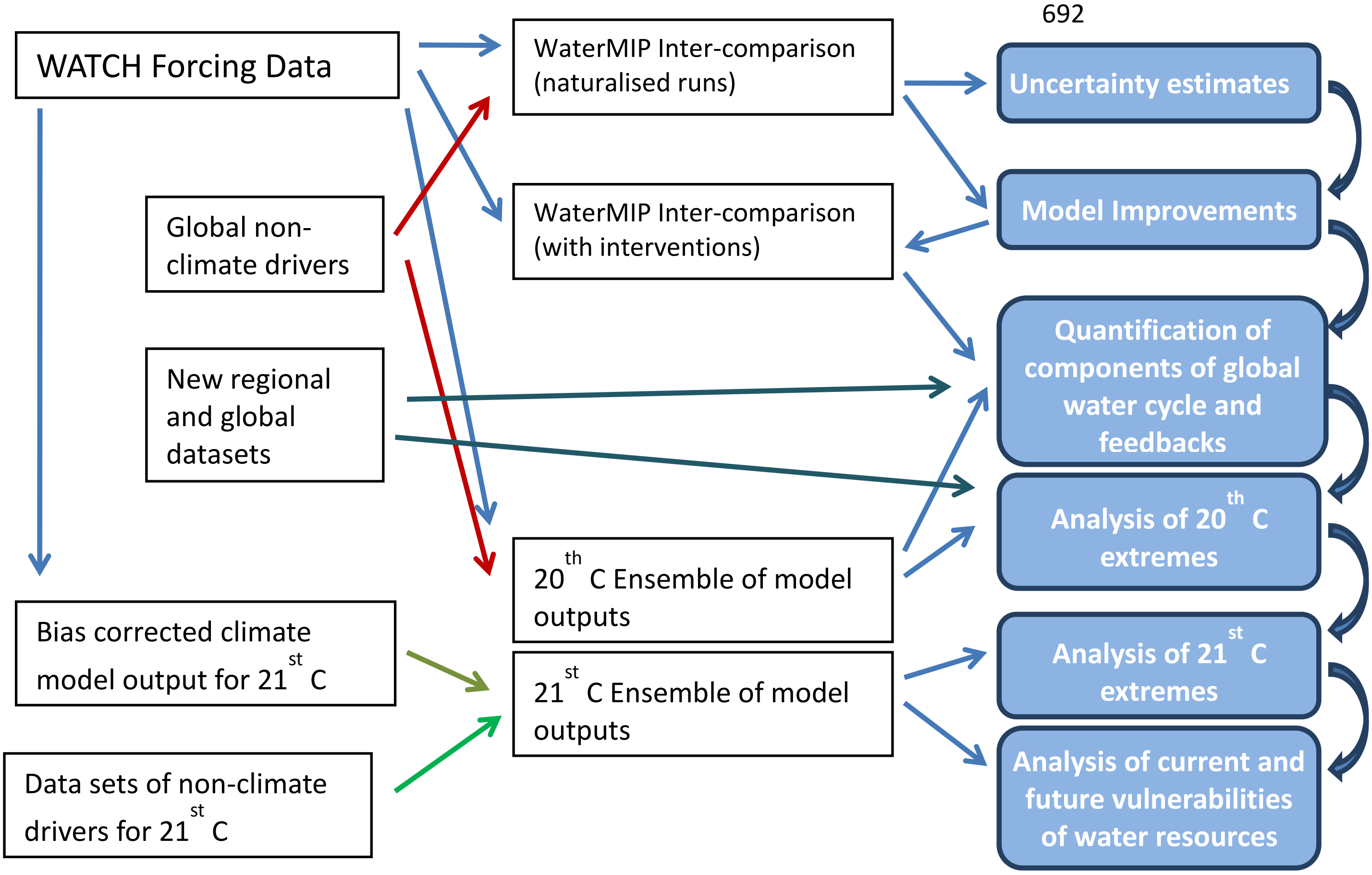

Figure 1. Analysis scheme of the WaterMIP intercomparison.

Figure 2. Time series of precipitation inputs, discharge and model runs, as described in text.

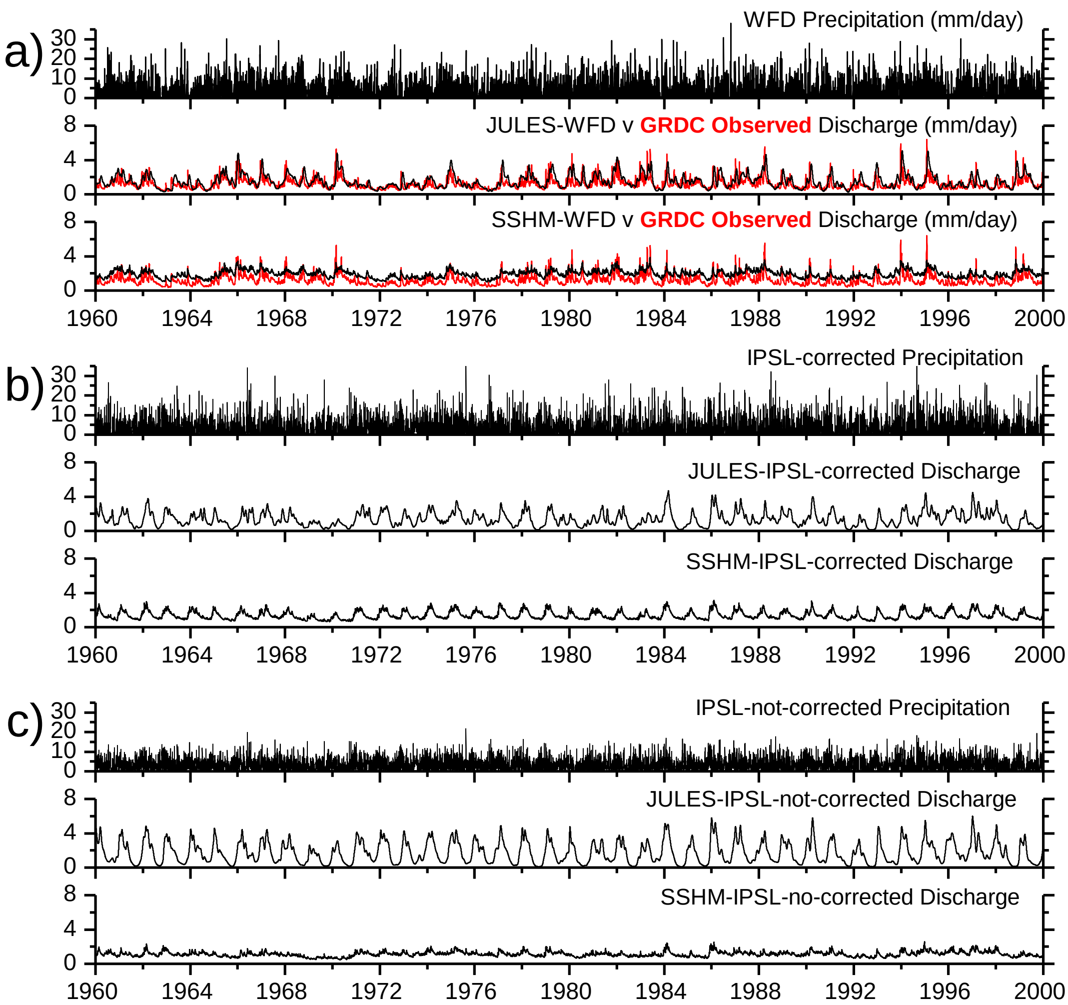

Figure 3. Mean Bias Error (MBE), annual amplitude and phase for model runs. Bars indicate the 95% confidence interval.